PRODUCTION SPECIFICATIONS – TAX DIVISION

Modifications:

Any modifications or deviations from these Production Specifications may only occur upon agreement between the parties. Any proposed production in a format other than those identified below, the proposed use of predictive coding, computer-assisted review, email threading or the use of de-duplication during the processing of documents, must be discussed with and approved by the legal and technical staff of the Tax Division and the methodology must be disclosed in the cover letter.

Production Format – ESI and Imaged Hard Copy:

Email and native document productions should include native file links, and all native documents and emails should be produced in their native format. All scanned paper shall be converted or processed to TIFF files (or color JPEGs when warranted), Bates numbered, and include fully searchable text as separate .txt files.

All documents must contain a Bates number – a unique, consistently formatted identifier of an alpha prefix and a fixed length number, i.e. ABC000001. This format shall remain consistent throughout the course of discovery. File names of native files should reflect the production Bates number.

All original hard copy documents shall reflect accurate document unitization including all attachments and container information. Unitization refers to identifying and marking the boundaries of documents within the collection, where a document is defined as the smallest physical fastened unit within a bundle. (e.g., staples, paperclips, rubber bands, folders, or tabs in a binder.)

Image File Format

- Images should be single-page, Group IV TIFF files, scanned at 300 dpi;
- File names must not contain embedded spaces;
- Bates numbers should be endorsed on the lower right corner of all images;
  - Any Confidential stamps or court ordered messages should be stamped on the lower left corner, but shall not impede the Bates number;
- Render to images photographs, AUTOCAD, Excel files, and QuickBooks database files:
  - AUTOCAD and photographs – scanned or processed to JPEG file format;
  - Excel spreadsheets should be produced in native file format with a single-page "PLACEHOLDER" image bearing the production Bates number. Applicable metadata should be produced in the DAT file;
  - QuickBooks database files should be produced in native .QBB or .QBW file format with a single-page "PLACEHOLDER" image bearing the production Bates number.

PRODUCTION SPECIFICATIONS – TAX DIVISION

- ▪ Images should be accompanied by a Concordance Image File (Opticon or log) that associates each Bates number to the image file path for each Bates numbered page in a page level comma-delimited file containing seven fields per line: PageID,VolumeLabel,ImageFilePath,DocumentBreak,FolderBreak,BoxBreak,PageCount

Metadata File Format – Concordance Data Load File Format (DAT)

Provide a data file (.DAT) containing fielded information to be loaded into a database for all documents in a production. While hard copy documents may not have fielded information (i.e. metadata), these documents should also be referenced in the .DAT file with as much fielded information as reasonably possible, such as DocID End and Extracted Text path.

The first line of a .DAT file must contain a header row identifying field names. The .DAT file must use the standard Concordance delimiters:

Comma	¶	ASCII Character (020)
Quote	þ	ASCII Character (254)
Newline	®	ASCII Character (174)
Multi-value Separator	;	ASCII Character (059)

Date fields should be provided in the format MM/DD/YYYYY, and should be the same format throughout the field. Attachments shall sequentially follow the parent document or email and their relationship should be identified using Parent ID and Child ID fields. All metadata associated with native documents, emails, and media files must be produced. See **Appendix A** for a list of basic metadata fields requested for production. This list is not exhaustive and requested fields may differ depending on the specifics of a particular matter. If producing fields outside of this list, provide a definition grid detailing the function of the field.

Extracted and OCR Text

Extracted text shall be produced as plain ASCII text files and produced at the document level. In the instance that extracted text is not available, such as produced documents with redactions, hard copy scanned files, or image-only files, OCR text shall be provided as plain ASCII text files and produced at the document level. A link to the text should be included in the .DAT file, but the text should not be embedded in the .DAT file.

Rev. 2019-08-07

PRODUCTION SPECIFICATIONS – TAX DIVISION

Electronic Document Production Specifications:

Any data or files that are encrypted to password protected should be decrypted and/or passwords removed prior to production. If a password cannot be removed, the password must be provided.

All hidden text such as track changes, hidden columns, etc., shall be expanded and rendered in the image file. For files that cannot be expanded the native files shall be produced with the image files.

Emails and other documents with embedded files should be provided in a way to attach the embedded document to the parent native file.

Spreadsheets and presentation files shall be produced in native document format.

Microsoft Access Databases or other Structured Data shall be produced in native MS Access 2010 or earlier format (.mdb) with a corresponding single-page "PLACEHOLDER" TIFF image, accompanied by a text load .DAT file containing Bates numbers, applicable metadata, and native and extracted text link. Please consult with the Tax Division's legal and tech department prior to submitting other database format.

Digital photographs shall be submitted as single-page JPEG files with a resolution equivalent to the original image as it was captured/created. All associated metadata and text fields shall be produced as well. All digital videos shall be submitted as native files with associated metadata and an accompanying single page "PLACEHOLDER" TIFF image.

Production Method

Data may be produced on CD, DVD, thumb drive, or hard drive; use the media requiring the least number of deliverables. The use of FTP sites or other cloud-based or internet-based data transfer sites is not permitted.

Media must be labeled with the containing at the minimum case/investigation name, case number (if applicable), production date, Bates range, and disc number.

The production shall be free of computer viruses. Any files found to include a virus shall be quarantined by the producing party and noted in a log to be provided in the production.

If the production should be encrypted or secured, the preferred format is to provide the media with self-extracting encrypted files and delivery of the associated password in a timely manner. Self-encrypting or keypad coded hard drives may be used in lieu of software encryption.

Rev. 2019-08-07

PRODUCTION SPECIFICATIONS – TAX DIVISION

Appendix A

An "X" indicates that the field should be populated in the load file produced whenever present for the appropriate record type. "Other ESI" includes non-email files, such as, but not limited to MS Office files, PDFs, and QuickBooks files.

Field Name	Field Description	Field Type	Hard Copy	Email	Other ESI	Calendar Items
DOCID	Must equal the value appearing in the BEGDOC# field and be UNIQUE.	Note Text	X	X	X	X
BEGDOC#	Start Bates (including prefix) -- No spaces or special characters.	Note Text	X	X	X	X
ENDDOC#	End Bates (including prefix) -- No spaces or special characters.	Note Text	X	X	X	X
COMPANIES	Company name of producing party (not law firm or vendor name).	Multi-Entry	X	X	X	X
CUSTODIAN	Individual Custodian Name/Shared Resource Name --format: Last, First or ABC Dept. Use consistent naming and formatting across all productions.	Multi-Entry	X	X	X	X
VOLUMENAME	Production volume number (e.g., ABC001-001).	Note Text	X	X	X	X
FAMILYRANGE	Range of the BEGDOC# value of the parent record to the ENDDOC# value (including prefix) of the last child record (for example, ABC-JD-00001201 - ABC-JD-00001220); populated for all documents in the group. Populate with BEGDOC# and ENDDOC# if the record is NOT in a family grouping.	Note Text	X	X	X	X
PARENTID	Parent record's BEGDOC#, including prefix (populated ONLY in child records).	Note Text	X	X	X	X
ATTACHMENTIDS	Child document list: BEGDOC# of each child (populated ONLY in parent records).	Multi-Entry	X	X	X	X
NUMPAGES	Page count.	Integer	X	X	X	X
EPROPERTIES	Indicate all that apply: Record Type: Edoc, Edoc Attachment, Email, Email Attachment, Hard Copy, Calendar Appt Other Notations: Translation, Translated Privilege Notations: Redacted, Privileged, Family Member of Priv Doc	Multi-Entry	X	X	X	X
TIMEZONE	The Time Zone in which the custodian is normally located.	Note Text		X	X	X
FOLDERLABEL	Email folder path or Hard Copy folder/binder title/label prepended with Custodian Name (sample: Smith, James-Inbox\Active).	Multi-Entry	X	X		X
FROM	Author of the Email or Calendar item (as formatted on the original).	Note Text		X		X
TO	Recipients of the Email or Calendar Item (as formatted on the original).	Multi-Entry		X		X
CC	Copyees of the Email or Calendar Item (as formatted on the original).	Multi-Entry		X		X
BCC	Blind Copyees of the Email or Calendar Item (as formatted on the original).	Multi-Entry		X		X
SUBJECT	Email or calendar item subject.	Note Text		X		X

Rev. 2019-08-07

PRODUCTION SPECIFICATIONS – TAX DIVISION

Field Name	Field Description	Field Type	Hard Copy	Email	Other ESI	Calendar Items
DATE_HC	Date of hard copy documents, if coded. Format: MM/DD/YYYY.	Date	X			
DOCDATE	Aggregate date field. Populate with: DATESAVED for Edocs; DATESENT for Emails; DATEAPPTSTART for calendar appointments; or DATE_HC for hard copy documents, if available. Format: MM/DD/YYYY.	Date	X	X	X	X
DATESENT	Date the Email or Calendar Item was sent. Format: MM/DD/YYYY.	Date		X		X
TIMESENT	Time Email or Calendar Item was sent -- Format: HH:MM:SS (use 24 hour times, e.g., 13:32:00 for 1:32 pm; time zone indicators cannot be included).	Time		X		X
DATERECEIVED	Date Email or Calendar Item was received. Format: MM/DD/YYYY.	Date		X		X
TIMERECEIVED	Time Email or Calendar Item was received. Format: HH:MM:SS (use 24 hour times, e.g., 13:32:00 for 1:32 pm; time zone indicators cannot be included).	Time		X		
HEADER	The internet header information from each Email.	Note Text		X		X
INTERNETMSGID	Globally unique identifier for a message which typically includes messageid and a domain name. Example: <0E6648D558F338179524D555@m1p.innovy.net	Note Text		X		X
MESSAGEID	Unique system identification number for the email message assigned by the proprietary email database/mailstore/post office file associated with centrally managed enterprise email servers. EntryID for Microsoft Outlook, the UniqueID (UNID) for Lotus Notes, or equivalent value for other proprietary mailstore formats.	Note Text		X		X
INREPLYTOID	Internet message ID of the Email replied to.	Note Text		X		X
CONVERSATIONINDEX	Email Thread Identification generated by the email application (Outlook).	Note Text		X		X
IMPORTANCE	Email flag indicating priority level set for message.	Note Text		X		X
DELIVRECEIPT	Delivery receipt request notification for Email messages.	Note Text		X		X
READRECEIPT	Read Receipt request notification for Email messages.	Note Text		X		X
SENSITIVITY	Sensitivity field from Email messages.	Note Text		X		X
REVISION	Revision number extracted from metadata of native file.	Note Text			X	
DATECREATED	Date electronic file was created. Format: MM/DD/YYYY.	Date			X	
TIMECREATED	Time electronic file was created. Format: HH:MM:SS (use 24 hour times, e.g., 13:32:00 for 1:32 pm; time zone indicators cannot be included).	Time			X	
DATESAVED	Date electronic file was last modified. Format: MM/DD/YYYY.	Date			X	
TIMESAVED	Time electronic file was last modified. Format: HH:MM:SS (use 24 hour times, e.g., 13:32:00 for 1:32 pm; time zone indicators cannot be included).	Time			X	
EORGANIZATION	"Company" field extracted from the metadata of a native file.	Note Text			X	
EAUTHOR	Author field value extracted from the metadata of a native file.	Note Text			X	

Page 5 of 7

PRODUCTION SPECIFICATIONS – TAX DIVISION

Field Name	Field Description	Field Type	Hard Copy	Email	Other ESI	Calendar Items
LAST_AUTHOR	Last Saved By field value extracted from metadata of a native file.	Note Text			X	
ESUBJECT	Subject field value extracted from metadata of a native file.	Note Text			X	
FILESIZE	File size in Bytes (integer value only - do not include unit of measure, thousands character, or decimal places - e.g., 1008 not 1,008).	Integer		X	X	X
FILENAME	File name of native file (Edocs or attachments to Email).	Note Text		X	X	X
APPLICATION	Application used to create native file (e.g., Excel, Outlook, Word).	Note Text		X	X	X
FILEEXTENSION	File extension of native file.	Max 5 chars		X	X	X
FILEPATH	File path to native file as it existed in original environment, prepended with Custodian Name; e.g. Smith, James-C\My Documents\Sales Info\ACME	Multi-Entry		X	X	X
DOCLINK	File path location to the current native file location on the delivery medium.	Note Text			X	
DATEAPPTSTART	Start date of calendar appointment. Format: MM/DD/YYYY.	Date				X
TIMEAPPTSTART	Start time of calendar appointment. Format: HH:MM:SS (use 24 hour times, e.g., 13:32:00 for 1:32 pm; time zone indicators cannot be included).	Time				X
DATEAPPTEND	End date of calendar appointment. Format: MM/DD/YYYY.	Date				X
TIMEAPPTEND	End time of calendar appointment. Format: HH:MM:SS (use 24 hour times, e.g., 13:32:00 for 1:32 pm; time zone indicators cannot be included).	Time				X
HASHMD5	Document MD5 hash value (used for deduplication or other processing).	Note Text		X	X	X
HASHSHA	Document SHA1 hash value (used for deduplication or other processing).	Note Text		X	X	X
SPEC#	Subpoena/request paragraph number to which the document is responsive.	Multi-Entry	X	X	X	X
SEARCHVALUES	Specific search terms used to identify the record as responsive (if used).	Multi-Entry	X	X	X	X

Additional Details about Family Relationships:

Attachments

     Graphic objects embedded in emails such as logos, letterheads, or backgrounds must remain as part of the email and not be extracted as separate attached documents.

     Pay special attention to the PARENTID, ATTACHMENTIDS, and FAMILYRANGE fields, which are used to track email families. While the example below relates to email families, all attachment relationships for all responsive documents are to be produced in this format.

PRODUCTION SPECIFICATIONS – TAX DIVISION

EXAMPLE: Consider ABC-JD-00000001 a 10-page parent email, with records ABC-JD-00000011 to ABC-JD-00000015, ABC-JD-00000016 to ABC-JD-00000020, and ABC-JD-00000021 to ABC-JD-00000025 as its attachments. Fields must be populated exactly as follows using the semicolon as the multi-entry delimiter for ATTACHMENTIDS:

BEGDOC#	ENDDOC#	PARENTID	ATTACHMENTIDS	FAMILYRANGE
ABC-JD-00000001	ABC-JD-00000010		ABC-JD-00000011; ABC-JD-00000016; ABC-JD-00000021	ABC-JD-00000001 – ABC-JD-00000025
ABC-JD-00000011	ABC-JD-00000015	ABC-JD-00000001		ABC-JD-00000001 – ABC-JD-00000025
ABC-JD-00000016	ABC-JD-00000020	ABC-JD-00000001		ABC-JD-00000001 – ABC-JD-00000025
ABC-JD-00000021	ABC-JD-00000025	ABC-JD-00000001		ABC-JD-00000001 – ABC-JD-00000025

Rev. 2019-08-07